



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/033,199	12/28/2001	Shivnath Babu	BABU 1-10-42	8231

47394 7590 02/14/2006

HITT GAINES, PC
LUCENT TECHNOLOGIES INC.
PO BOX 832570
RICHARDSON, TX 75083

EXAMINER

LERNER, MARTIN

ART UNIT	PAPER NUMBER
----------	--------------

2654

DATE MAILED: 02/14/2006

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No. 10/033,199	Applicant(s) BABU ET AL.	
	Examiner Martin Lerner	Art Unit 2654	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) ☒ Responsive to communication(s) filed on 19 December 2005.
- 2a) ☒ This action is **FINAL**. 2b) ☐ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) ☒ Claim(s) 1 to 24 is/are pending in the application.
- 4a) Of the above claim(s) _____ is/are withdrawn from consideration.
- 5) ☐ Claim(s) _____ is/are allowed.
- 6) ☒ Claim(s) 1 to 22 and 24 is/are rejected.
- 7) ☒ Claim(s) 23 is/are objected to.
- 8) ☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☒ The drawing(s) filed on 19 December 2005 is/are: a) ☒ accepted or b) ☐ objected to by the Examiner.
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some * c) ☐ None of:
1. ☐ Certified copies of the priority documents have been received.
 2. ☐ Certified copies of the priority documents have been received in Application No. _____.
 3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

- | | |
|--|---|
| 1) <input type="checkbox"/> Notice of References Cited (PTO-892) | 4) <input type="checkbox"/> Interview Summary (PTO-413)
Paper No(s)/Mail Date. _____ |
| 2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948) | 5) <input type="checkbox"/> Notice of Informal Patent Application (PTO-152) |
| 3) <input type="checkbox"/> Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08)
Paper No(s)/Mail Date _____ | 6) <input type="checkbox"/> Other: _____ |

DETAILED ACTION

Claim Rejections - 35 USC § 102

1. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless –

(e) the invention was described in (1) an application for patent, published under section 122(b), by another filed in the United States before the invention by the applicant for patent or (2) a patent granted on an application for patent by another filed in the United States before the invention by the applicant for patent, except that an international application filed under the treaty defined in section 351(a) shall have the effects for purposes of this subsection of an application filed in the United States only if the international application designated the United States and was published under Article 21(2) of such treaty in the English language.

2. Claims 1, 9, and 17 are rejected under 35 U.S.C. 102(e) as being anticipated by *Fayyad et al.* ('882).

Fayyad et al. ('882) discloses a system and method for database management, comprising:

“a table modeler that discovers at least one model of data mining models with guaranteed error bounds of at least one attribute in said data table in terms of other attributes in different columns of said data table” – the invention evaluates a database 10 having many records stored on storage devices; each record in the database 10 has many attributes or fields which for a representative database might include age, income, number of years of employment, census data, etc. (column 4, line 60 to column 5, line 2); implicitly, a plurality of records, where each record has a number of attributes, is a table; a data clustering model (“table modeler”) is produced that implements a data

mining engine for answering queries about data records in the database (column 5, lines 20 to 25); accuracy parameters ("guaranteed error bounds") are used to control the clustering; an accuracy parameter can be the percentage by which the number of points is allowed to deviate from an expected value or the probability of a tile satisfying the accuracy criterion (column 9, line 63 to column 10, line 42); Table 1 shows age, salary, and years employed as "different columns" of a data table (column 5, lines 25 to 39);

"a model selector, associated with said table modeler, that selects a subset of said at least one model to form a basis upon which to compress said data table to form a compressed data table" – a data mining engine 12 forms conclusions about the accuracy of an initial model (M), and the model is refined until the model more accurately represents the data stored in the database (column 9, lines 37 to 62); a cluster must satisfy an accuracy requirement for the model to be judged suitable (column 10, lines 33 to 42); a model represents a compressed version of records in data database 10 (Abstract); a model is formed by selecting "a subset of said at least one model" at least because outlier data points, which have distances greater than a constant ξ for a cluster, are not members of clusters if a specified memory condition is exceeded (column 18, line 25 to column 19, line 13).

Claim Rejections - 35 USC § 103

3. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

Art Unit: 2654

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

4. Claims 2, 4, 5, 7, 8, 10, 12, 13, 16, 18, 20, 21, and 24 are rejected under 35 U.S.C. 103(a) as being unpatentable over *Fayyad et al.* ('882) in view of *Agrawal* ('311).

Concerning claims 2, 10, and 18, *Fayyad et al.* ('882) does not disclose specifics about the modeling process as employing classification and regression tree (CaRT) data mining to model attributes. However, *Agrawal* ('311) suggests data mining with decision trees for modeling records having one or more attribute values may be by classification and regression trees. (Column 5, Line 63 to Column 6, Line 7; Column 6, Lines 58 to 67) The stated objective is provide an efficient method for generating a decision-tree classifier that is compact, accurate, has short training times, and is scalable. (Column 3, Lines 11 to 24) It would have been obvious to one having ordinary skill in the art to employ classification and regression trees for data mining of model attributes as taught by *Agrawal* ('311) in the multi-dimensional database record compression of *Fayyad et al.* ('882) for the purpose of generating decision trees by a classifier that is compact, accurate, has short training times, and is scalable.

Concerning claims 4, 12, and 20, *Agrawal* ('311) discloses pruning for short training time (column 8, line 40 ff).

Concerning claims 5, 13, and 21, *Agrawal* ('311) discloses pruning for representing misclassification errors based upon encoding costs (column 9, lines 34 to 54), which is equivalent to a "scoring-based method".

Concerning claim 7, *Agrawal* ('311) discloses data mining with decision trees for modeling records having one or more attribute values may be by classification and regression trees (column 5, line 63 to column 6, line 7; column 6, lines 58 to 67); implicitly, values of attributes are stored as models and not as data points, so values "are not explicitly stored therein."

Concerning claims 8, 16, and 24, *Agrawal* ('311) discloses a greedy algorithm may be used for subsetting (column 8, line 3).

5. Claims 2, 3, 10, 11, 18, and 19 are rejected under 35 U.S.C. 103(a) as being unpatentable over *Fayyad et al.* ('882) in view of *Pednault*.

Fayyad et al. ('882) does not disclose specifics about the modeling process as employing classification and regression trees or a Bayesian network. However, *Pednault* teaches a method for constructing predictive models that involve Bayesian networks (column 2, lines 20 to 30 and column 2, lines 45 to 52) and classification and regression trees (column 2, lines 35 to 45). The objective is to provide a method of handling missing values. It would have been obvious to one having ordinary skill in the art to employ classification and regression trees or Bayesian networks as suggested by *Pednault* in the multi-dimensional database record compression of *Fayyad et al.* ('882) for the purpose of providing a method for handling missing values.

6. Claims 6, 14, and 22 are rejected under 35 U.S.C. 103(a) as being unpatentable over *Fayyad et al.* ('882) in view of *Chakrabarti et al.* ('005).

Fayyad et al. ('882) omits selecting a subset based upon a compression ratio. However, *Chakrabarti et al.* ('005) teaches a method for data mining where a compression ratio is an indicator of complexity of compressed files. (Column 16, Lines 18 to 25) The objective is to select candidate data patterns from a dataset based on the variations of support values of a pattern. (Column 5, Lines 4 to 14) It would have been obvious to one having ordinary skill in the art to select a data subset based upon a compression ratio as suggested by *Chakrabarti et al.* ('005) in the multi-dimensional database record compression of *Fayyad et al.* ('882) for the purpose of selecting candidate data patterns from a dataset.

7. Claim 15 is rejected under 35 U.S.C. 103(a) as being unpatentable over *Fayyad et al.* ('882) in view of *Agrawal et al.* ('048).

Fayyad et al. ('882) does not disclose that a process by which a model selector selects a subset is NP-hard. However, *Agrawal et al.* ('048) teaches that, in general, an optimized rule mining problem is NP-hard. (Column 4, Lines 9 to 14) The objective is to provide a method for identifying database association rules which are optimal at upper and lower support-confidence borders. (Column 4, Line 30 to Column 5, Line 45) It would have been obvious to one having ordinary skill in the art that model selection is an NP-hard algorithm as suggested by *Agrawal et al.* ('048) in the multi-dimensional database record compression of *Fayyad et al.* ('882) for the purpose of providing optimal association rules at upper and lower support-confidence borders.

Allowable Subject Matter

8. Claim 23 is objected to as being dependent upon a rejected base claim, but would be allowable if rewritten in independent form including all of the limitations of the base claim and any intervening claims.

Response to Arguments

9. Applicants' arguments filed 19 December 2005 have been fully considered but they are not persuasive.

Applicants argue that *Fayyad et al.* ('882) fails to anticipate independent claims 1, 9, and 17 under 35 U.S.C. §102(e) because the reference does not teach discovering at least one model of data mining models with guaranteed error bounds of at least one attribute in a data table in terms of other attributes in different columns of the data table. Specifically, Applicants say that *Fayyad et al.* ('882) discloses employing a data mining engine to produce a clustering model derived from a database, and that accuracy parameters are used to control the clustering initialization process. However, Applicants maintain that the accuracy parameters are not guaranteed error bounds of an attribute of the database in terms of other attributes in different columns of the database. Applicants note that, instead, *Fayyad et al.* ('882) teaches the accuracy parameters are adjustable parameters used to determine, for example, the number of data points per attribute partition or the probability of a partition satisfying an accuracy criterion. This position is not convincing.

At the outset, it is noted that Applicants' term "guaranteed error bounds" is not a term of art, and is not expressly defined by Applicants' Specification. Accordingly, the term "guaranteed error bounds" should be given a broadest reasonable interpretation that is not inconsistent with the Specification. During patent examination, the pending claims must be "given their broadest reasonable interpretation consistent with the specification." *In re Hyatt*, 211 F.3d 1367, 1372, 54 USPQ2d 1664, 1667 (Fed. Cir. 2000). Applicant always has the opportunity to amend the claims during prosecution, and broad interpretation by the examiner reduces the possibility that the claim, once issued, will be interpreted more broadly than is justified. *In re Prater*, 415 F.2d 1393, 1404-05, 162 USPQ 541, 550-51 (CCPA 1969) See MPEP 2111. The prior art need not expressly disclose a term employed by Applicants, as it is always possible that the prior art can disclose an equivalent term by saying the same thing in different words.

Generally, Applicants' Specification discloses that the term "guaranteed error bounds" relates to how well a model represents the data after compression. By its nature, compression is lossy and reduces the accuracy of the model with respect the original data. Thus, Applicants' "guaranteed error bounds" describe a range of error produced by a model when a model is employed instead of the original data. The less accurate a model is, the more error is produced by a model.

Fayyad et al. ('882) discloses "a number of accuracy parameters" (column 9, line 64), including a number of points per attribute partition or tile (column 10, lines 1 to 2). One skilled in the art would recognize that the number of points per tile directly affects the accuracy of the model. If there are more points per tile, then each tile is less

accurate with respect to the original data. The model, as a collection of tiles, is then more granular and less accurate. By contrast, if there are fewer points per tile, then each tile is more accurate with respect to the original data. However, fewer points per tile implies that the model is less compressed. A model that is less compressed involves greater computational complexity.

Specifically, *Fayyad et al.* ('882) discloses a TileAccuracy as a percentile value that is equivalent to Applicants' "guaranteed error bounds". A TileAccuracy of 80% would mean that for a tile to be judged as accurate, the number of data points falling within the tile must be above or below the model prediction by no more than 20%. (Column 10, Lines 13 to 25) Similarly, the TilePercentage gives a percentage of tiles within a cluster that must satisfy the accuracy requirement for the model to be judged suitable. (Column 10, Lines 33 to 37) Also, *Fayyad et al.* ('882) describes a "confidence interval" of a TilePercentage (column 10, lines 51 to 66). Furthermore, *Fayyad et al.* ('882) says that judging an accuracy of a model involves determining a maximum positive error tile and a maximum negative error tile. If there is a maximum positive error tile not satisfying the accuracy criterion, then clustering must be reformulated to satisfy the accuracy criterion. (Column 14, Lines 30 to 35; Column 22, Lines 40 to 55)

Thus, *Fayyad et al.* ('882) discloses several formulations that are equivalent to Applicants' term "guaranteed error bounds". *Fayyad et al.* ('882) provides a percentile accuracy value of 80% can be required for a model to be judged accurate. There are "confidence intervals" for a model represented by a Gaussian distribution to control the

accuracy during a query process. Also, there are maximum errors for tiles to satisfy an accuracy criterion. All of these formulations disclosed by *Fayyad et al.* ('882) are equivalent to Applicants' "guaranteed error bounds". Admittedly, *Fayyad et al.* ('882) does not expressly disclose the term "guaranteed error bounds". However, Applicants are simply saying the same thing in different words. Thus, the term "guaranteed error bounds" is equivalent to *Fayyad et al.* ('882)'s "percentile tile accuracy", "confidence intervals", and "maximum positive tile error" under principles of broadest reasonable interpretation.

Finally, it is noted that each point of *Fayyad et al.* ('882) represents one attribute in terms of another attribute (e.g., salary versus years). Similarly, each tile of *Fayyad et al.* ('882) represents one attribute in terms of another attribute in a compressed model. A record containing many attributes or fields presents each attribute as a column of a table. (Table 1: Column 5, Lines 30 to 39) Thus, *Fayyad et al.* ('882)'s "percentile tile accuracy", "confidence intervals", and "maximum positive tile error" represent how well a model represents guaranteed error bounds of an attribute of a database in terms of other attributes in different columns of the database.

Therefore, the rejections of claims 1, 9, and 17 under 35 U.S.C. §102(e) as being anticipated by *Fayyad et al.* ('882), of claims 2, 4, 5, 7, 8, 10, 12, 13, 16, 18, 20, 21, and 24 under 35 U.S.C. §103(a) as being unpatentable over *Fayyad et al.* ('882) in view of *Agrawal* ('311), of claims 2, 3, 10, 11, 18, and 19 under 35 U.S.C. §103(a) as being unpatentable over *Fayyad et al.* ('882) in view of *Pednault*, of claims 6, 14, and 22 under 35 U.S.C. §103(a) as being unpatentable over *Fayyad et al.* ('882) in view of

Chakrabarti et al. ('005), and of claim 15 under 35 U.S.C. §103(a) as being unpatentable over *Fayyad et al.* ('882) in view of *Agrawal et al.* ('048), are proper.

Conclusion

10. **THIS ACTION IS MADE FINAL.** Applicants are reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the mailing date of this final action.

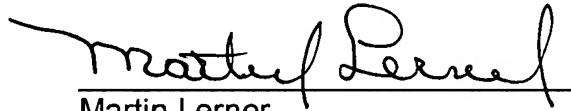
Any inquiry concerning this communication or earlier communications from the examiner should be directed to Martin Lerner whose telephone number is (571) 272-7608. The examiner can normally be reached on 8:30 AM to 6:00 PM Monday to Thursday.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, David R. Hudspeth can be reached on (571) 272-7843. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Art Unit: 2654

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

ML
2/7/06



Martin Lerner
Examiner
Group Art Unit 2654